Processing personal data without the consent of the data subject for the development and use of language resources

Kelli, Aleksei

Linköping University Electronic Press
2019-05-28

þÿ K e l l i , A , L i n d é n , K , V i d e r , K , K a m o c k i , P , B i r a t o n a s , R , C a l a m a Gavriilidou , M & Stranák , P 2019 , Processing personal data without the consent of the data subject for the development and use of language resources . in I Skadina & M Eskevich (eds) , Selected papers from the CLARIN Annual Conference 2018, Pisa, 8-10 October 2018 . Linköping Electronic Conference Proceedings , no. 159 , Linköping University Electronic Press , Linköping , pp. 72-82 , CLARIN Annual Conference , Pisa , Italy , 08/10/2018 . < https://www.ep.liu.se/ecp/article.asp?issue=159&article=008&volume=0 >

http://hdl.handle.net/10138/316219

Processing personal data without the consent of the data subject for the development and use of language resources

Aleksei Kelli
University of Tartu
Estonia
aleksei.kelli@ut.ee

Krister Lindén
University of Helsinki
Finland
krister.linden@helsinki.fi

Kadri Vider
University of Tartu
Estonia
kadri.vider@ut.ee

Pawel Kamocki
ELDA, France /
IDS Mannheim,
Germany
pawel.kamocki@gmail.com

Ramūnas Birštonas
Vilnius University
Lithuania
ramunas.birstonas@tf.vu.lt

Silvia Calamai
University of Siena
Italy
silvia.calamai@unisi.it

Penny Labropoulou
ILSP/ARC, Greece
penny@ilsp.gr

Maria Gavrilidou
ILSP/ARC, Greece
maria@ilsp.gr

Pavel Straňák
Charles University, Czechia
stranak@ufal.mff.cuni.cz

Abstract

The development and use of language resources often involve the processing of personal data. The General Data Protection Regulation (GDPR) establishes an EU-wide framework for the processing of personal data for research purposes while at the same time allowing for some flexibility on the part of the Member States. The paper discusses the legal framework for language research following the entry into force of the GDPR. In the first section, we present some fundamental concepts of data protection relevant to language research. In the second section, the general framework of processing personal data for research purposes is discussed. In the last section, we focus on the models that certain EU Member States use to regulate data processing for research purposes.

1 Introduction[1]

Language resources (LRs) contain material subject to various legal regimes. For instance, they may contain copyright protected works, objects of related rights (performances) and personal data. This affects the way language resources are collected and used. Intellectual property issues relating to language resources have been previously addressed (see Kelli et al. 2015). The focus of this article is on personal data protection. More precisely on the processing of personal data for research purposes without the data subject's consent within the framework of language research. Personal data issues are relevant for language resources, given that they potentially contain oral speech or written text which relates to a natural

Aleksei Kelli, Krister Lindén, Kadri Vider, Pawel Kamocki, Ramūnas Birštonas, Silvia Calamai, Penny Labropoulou, Maria Gavrilidou, and Pavel Straňák 2019. Processing personal data without the consent of the data subject for the development and use of language resources. *Selected papers from the CLARIN Annual Conference 2018*. Linköping Electronic Conference Proceedings 159: 72–82.

person.[2] In the CLARIN Virtual Language Observatory (VLO), approximately 95,502 language resources[3] could contain personal data.[4]

Although the General Data Protection Regulation[5] (GDPR) provides a general framework for personal data protection, it leaves a certain degree of freedom for the EU Member States to regulate the processing of personal data in different contexts (including research, see GDPR Art. 89 (3)). Even the duration of personal data protection is up to the Member States.[6] For instance, according to the Estonian Personal Data Protection Act, the data subject's rights are protected during the lifetime of the data subject and for ten years after the death of the data subject. In the case of minors, the duration is the lifetime and twenty years (§ 9).[7] This means that the Member States can adopt different regulatory models.

This article preliminarily maps the regulatory framework for processing personal data for research purposes. It also provides insights into different national models.[8] The picture is further complicated by the fact that, in addition to the GDPR and national laws directly related to data protection, other national legislation may add regulations to data protection and privacy in particular contexts, e.g. health care. Before concentrating on the data processing for research purposes, key concepts of the data protection framework are addressed.

2 Data subject, personal data and data processing

The data subject is defined through the concept of personal data. Personal data is "any information relating to an identified or identifiable natural person ('data subject')" (GDPR Art. 4). Publicly available personal data is also protectable (C-73/07). According to the Article 29 Working Party[9] (WP29), information contained in free text in an electronic document may qualify as personal data. It does not have to be in a structured database (2007: 8).

The identifiability is a crucial issue since data not relating to a natural person (incl. anonymous data) is not subject to the GDPR requirements (See GDPR Recital 26). One option to avoid problems with personal data protection is the anonymisation of data used for language research. However, it should be kept in mind that the process of rendering personal data anonymous is an instance of further processing which has to follow the data protection requirements (WP29 2014a: 3). It is also slightly complicated as combining already anonymised data sources may again make their data personal, and in some cases, anonymisation may render the data useless for research purposes. For other protective measures, see Section 3.2 below.

A natural person can be identified by reference to an identifier (e.g., name, identification number), location data and physiological, genetic, mental, economic, cultural or social information (GDPR Art. 4). According to WP29 sound and image data qualify as personal data insofar as they may represent information on an individual (WP29 2007: 7). It means that LRs containing oral speech are subject to the GDPR. A question can be raised whether speech and voice as such constitute personal data where there is no additional information leading to a specific individual. It is a question related to identifiability. As suggested in the literature, data that are not identifiable for one person may be identifiable for another. Data can also become identifiable through combination with other data sets. Identifiability is a broad category depending on how much effort must be deemed 'reasonable' (Oostveen 2016: 306).

[2] For instance, according to the Court of Justice of the European Union (CJEU) the concept of personal data covers the name of a person (C-101/01).

[3] Resource type: Audio, Radio, Sound, Speech, Spontaneous, Television or Video.

[4] Language resources with written text may also contain personal data, but this is not as prominent as in the case of audio and/or visual material (e.g. interviews or photos of a certain person).

[5] The GDPR is applicable in all EU Member States from 25 May 2018. It replaces the Data Protection Directive.

[6] The GDPR does not apply to the personal data of deceased persons. Member States may establish the relevant regulation (GDPR Recital 27).

[7] The duration of personal data protection is rather complicated issue since the deceased person's data may still refer to a living person (WP29 2007: 22).

[8] For lack of space not all the EU countries are addressed in the present paper.

[9] According to the Data Protection Directive the Working Party on the Protection of Individuals with regard to the Processing of Personal Data (WP29) is composed of a representative of the supervisory authority or authorities designated by each Member State and of a representative of the authority or authorities established for the Community institutions and bodies, and of a representative of the Commission.

Voice can be considered biometric data (see González-Rodríguez et al. 2008; Jain et al. 2004).[10] Biometric data for uniquely identifying a natural person belongs to a special category of personal data[11] the processing of which is even more restricted than for other personal data. A similar case is that of photos depicting people. Here the GDPR provides a clarification: "The processing of photographs should not systematically be considered to be processing of special categories of personal data as they are covered by the definition of biometric data only when processed through a specific technical means allowing the unique identification or authentication of a natural person" (Recital 51). This should be applicable in case of speech and video as well. Therefore, the requirements concerning the processing of special categories of personal data apply in case oral speech contained in language resources is used for the identification of natural persons.

The GDPR defines processing very broadly. It includes, among other things, collection, structuring, storage, adaptation, use, making available or destruction (GDPR Art. 4). It means that the development and use of LRs containing personal data constitutes processing.

Personal data protection requirements do not have to be followed in case the processing of personal data is done by a natural person in the course of a purely personal or household activity (GDPR Art. 2 (2)). It is debatable if the private use exemption is applicable for research as well.

3 Processing personal data for research purposes

3.1 General framework

The General Data Protection Regulation sets forth the following principles relating to processing of personal data (incl. for research purposes): 1) lawfulness, fairness and transparency; 2) purpose limitation (data is collected for specified, explicit and legitimate purposes); 3) data minimisation (the collection and use of data is as limited as possible); 4) accuracy; 5) storage limitation (kept for no longer than is necessary); 6) integrity and confidentiality; 7) accountability (Art. 5). It is explained that further processing for research is compatible with the initial purposes. Personal data can be stored for more extended periods for research purposes (GDPR Art. 5).

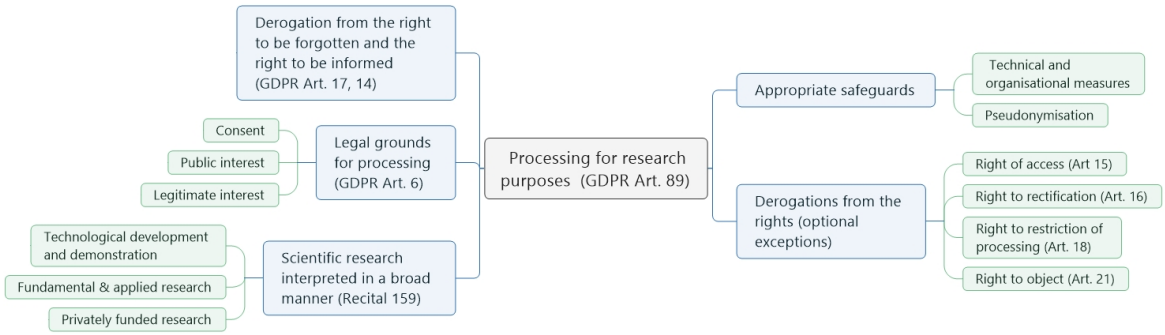

Figure 1: Processing of personal data for research purposes.

The GDPR provides six legal grounds for processing personal data: 1) consent; 2) performance of a contract; 3) compliance with a legal obligation; 4) protection of the vital interests; 5) the public interest or in the exercise of official authority; 6) legitimate interests (Art. 6).

As seen, the processing for research purposes is not an individual legal ground. Therefore, the processing for research purposes has to take place within the existing six grounds. The processing can rely on consent (for further discussion on consent see WP29 2017), the performance of a task carried out in the public interest or the legitimate interests.

[10] The GDPR defines biometric data as "personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data" (Art. 4).

[11] The GDPR defines special categories of personal data as "data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation".

It is not entirely clear when the processing for research purposes must rely on consent and when the public interest and legitimate interest can be used as grounds. Note, however, that consent is needed at least if the aim is to make personal data publicly available because public or legitimate interest require protective measures limiting access. Consent may also be mandated by national legislation in particular cases, e.g. when collecting health data.

It can be presumed that the processing based on the data subject's consent provides the highest protection of his/her fundamental rights (privacy, integrity, self-realisation). The data subject may even withdraw his/her consent without any legal consequences (GDPR Art. 7 (3)). The controller[12] has to be able to prove the existence of the consent (GDPR Art. 7 (1)). WP29 explains that consent "focuses on the self-determination of the data subject as a ground for legitimacy. All other grounds, in contrast, allow processing – subject to safeguards and measures – in situations where, irrespective of consent, it is appropriate and necessary to process the data within a certain context in pursuit of a specific legitimate interest" (2014: 13).

In case where the acquisition of consent is very complicated or administratively burdensome (e.g., anonymous web posts, legacy resources, public videos and so forth) the question arises which legal ground is relevant. According to WP29, the performance of a task carried out in the public interest is another ground for processing personal data in the research context (2014b: 21-23). The concept of research in the public interest[13] can usually be invoked by research projects affiliated with universities or research institutions having a legal mandate to do research in the public interest[14], i.e. agencies acting on behalf of a Member State.

The GDPR also names the legitimate interests as a legal ground for processing. The concept of legitimate interest is rather complicated and requires weighing different interests.[15] According to WP29, legitimate interest can serve as a legal ground for processing personal data in the research context (2014b: 24-25). The legitimate interest is most likely relevant for commercial research.

Before addressing specific requirements concerning the processing of personal data for research, it is necessary to outline the concept of research in the data protection context. The GDPR defines research broadly so that it covers "technological development and demonstration, fundamental research, applied research and privately funded research" (Recital 159).

The GDPR provides the following requirements for processing data for research purposes (Art. 89):

1. processing for research purposes is subject to appropriate safeguards. The safeguards ensure that technical and organisational measures are in place in particular to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner;
2. the Member States may limit the following data subject's rights for research purposes (optional exceptions):
    a) the right of access by the data subject (Art. 15);
    b) the right to rectification (Art. 16);
    c) the right to the restriction of processing (Art. 18);
    d) the right to object (Art. 21);

[12] The GDPR defines the controller as "the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data" (Art. 4 (7)).

[13] According to the GDPR, processing is lawful if it is "processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller" (Art. 6e).

[14] For instance, according to the Estonian Organisation of Research and Development Act (ORDA) a research and development institution is a legal person or an institution in the case of which the principal activity is carrying out basic research, applied research or development, or several of the aforementioned activities (§ 3 (1) clause 1).

[15] According to the GDPR, processing is lawful if it is "necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data" (Art. 6f).

There is also a mandatory exception[16] concerning the right to be forgotten[17] and right to be informed about the processing:

1. the right to be forgotten is limited to the extent that processing is necessary for research purposes in so far as the right to be forgotten is likely to render impossible or seriously impair the achievement of the objectives of that processing (GDPR Art. 17 (3)d);
2. the right to be informed about the processing of personal data is limited insofar as the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for research purposes and it is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases, the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available (GDPR Art. 14 (5) b).

Safeguards are described in the next section. The implementation of optional exceptions is outlined in the section dedicated to national models.

3.2 Appropriate safeguards

Protective measures may be of a technical or organisational nature. The technical measures may concern the data, medium or procedure, and the organisational measures may concern the staff, documentation or procedures. Examples of **technical measures** concerning 1) the *data* are pseudonymization, anonymization or aggregates of personal data; 2) the *medium* are encryption of personal data, internal measures by the data controller and data processor to prevent access to personal data, or measures to verify and prove who has registered, changed or transferred personal data; 3) the *procedure* are measures to continuously safeguard confidentiality, integrity, availability and resilience of processing systems and services in relation to the processing of personal data including the capacity to restore the availability or to safeguard access to personal data in a timely manner in the event of a physical or technical incidents.

Examples of **organisational measures** concerning 1) the *staff*: are appointing a data protection officer, or measures to raise the competence of the staff dealing with personal data, 2) the *documentation* are risk assessments, controller's record of processing activities, data processor agreements, guidelines, or non-disclosure agreements, 3) the *procedures* are a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing, establishing specific procedures to ensure that Union and Member State law are adhered to in case personal data is transferred or processed for some other purpose, or carrying out a data protection impact assessment.

4 National models

In **Czechia** application of the DGPR is still in progress. At the time of writing of this article, there is a mixed model of the previous Personal Data Protection Law (Czech law) https://zakonyprolidi.cz/cs/2000-101 and the GDPR regulation that overrides some parts of it. Parts that are not overridden by the regulation are still valid – e.g. existence and duties of the Office of Personal Data Protection established by law 101/2000 – until new "adaptation law" that replaces law 101/2000 is passed. The proposal of such a new law adopting the GDPR is now in the legislative process. There was a government proposal in March 2018, and after going through committees and debates in the Chamber of Deputies (lower chamber) of the parliament where it went through 29 amendments, it was passed to the Senate (upper chamber) on 8 January 2019. Currently, it is in the Senate committees, collecting more proposals for amendments. The proposal will be debated on the Senate floor on 30 January.[18] Several of the proposed amendments relate to research exceptions. At the time of passing the proposal to Senate, some deputies added § 16 that was not present in the government proposal. It is titled *"Collecting personal data for scientific or historical research or for statistical purposes":*

- Processing for these purposes is allowed provided that various protecting measures incl. pseudonymisation, maintaining processing logs according to Art. 5 of the GDPR, regular audits,

[16] Mandatory exceptions are directly applicable. They do not need to be incorporated into the national laws.

[17] The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her (GDPR Art. 17 (1)).

[18] The current status of the proposal and all suggested changes can be followed (in Czech) at the website of Czech Parliament: http://public.psp.cz/en/sqw/historie.sqw?o=8&T=138 (28.1.2019).

etc., are followed. The measures shall be "commensurate with state of the art, the cost of execution, the nature, scope, context and purposes of the processing."[19]

- § 16 ends with this sentence: "Article 15[20] and, to its corresponding extent, Article 5[21] of the GDPR […] shall not apply where processing is necessary for the purposes of scientific research, and the provision of information would require a disproportionate effort."

Thus, the current proposed law would allow scientific processing including large scale data collection for Natural Language Processing provided that best effort is taken to protect personal data. However, the version has to be adopted yet.

The **Estonian** Personal Data Protection Act (PDPA 2018a) sets the following requirements for the processing of personal data for scientific research (§ 6):

1) Personal data may be processed without the consent of the data subject for research purposes mainly if data has undergone pseudonymisation.
2) Processing of data without consent for scientific research in a format which enables identification of the data subject is permitted only if the following conditions are met:
   a) after removal of the data enabling identification, the goals of data processing would not be achievable, or achievement thereof would be unreasonably difficult;
   b) the person carrying out the scientific research finds that there is a predominant public interest for such processing;
   c) obligations of the data subject are not changed by the processed personal data, and the rights of the data subject are not excessively damaged in any other manner.
3) The data controller may limit the data subject's right of access, right to rectification, right to the restriction of processing and right to object in so far as the exercise of these rights are likely to render impossible or seriously impair the achievement of the objectives of the processing for research purposes.
4) In case of processing of special categories of personal data an ethics committee in the corresponding area verifies, before the commencement of the processing, compliance with the requirements set out in this section. In the absence of an ethics committee in a specific area, the Data Protection Authority verifies the fulfilment of requirements.

According to the **Finnish** model, the Data Protection Act (DPA 2018) and the preamble of the Government Proposal for Data Protection Act (Draft PDPA 2018b) outline the following conditions for processing personal data for scientific research:

1) **Data protection in general:** The legal basis for processing personal data by scientific researchers is, according to GDPR §6.1e, i.e. *performance of a task carried out in the public interest* based on the research organisation's legal mandate to do research as long as, according to GDPR §5.1f, the data is processed in a manner that ensures appropriate security of the personal data. Research organisations also have the right to store personal data as long as necessary and reuse them for secondary research purposes based on GDPR §5.1b, i.e. *further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall … not be considered to be incompatible with the initial purposes.*
2) **Data protection in special categories:** According to DPA 2018 §6.7 and §6.8, the above also applies to personal data for special categories mentioned in GDPR §9.1 (with the exception of archiving genetic data) provided that suitable and specific technical and organisational measures are implemented to safeguard the fundamental rights and the interests of the data subject.
3) **Limitations to the rights of data subjects:** According to DPA 2018 §31.1, the data subjects may have limited rights to stop processing of personal data for scientific and historical research if the processing is necessary for carrying out the research, in which case the motivation for why the processing is necessary should be included in a research plan identifying the Principal Investigator.
4) **Limitations to the rights of data subjects in special categories:** According to DPA 2018 §31.3, if limitations to the rights of the research subject are applied to personal data in special categories,

[19] http://www.psp.cz/sqw/text/tiskt.sqw?o=8&ct=138&ct1=0&v=PZ&pn=12&pt=1 (28.1.2019)
[20] Right of access by the data subject.
[21] Principles relating to processing of personal data.

the research plan should assess how the limitations impact the rights and freedoms of the data subject. The written assessment must be delivered to the Data Protection Ombudsman ahead of starting the processing.

In **France**, the national law completing the GDPR initially proposed on 13 December 2017, has finally been adopted on 20 June 2018. It has since been amended twice: by the Decree n° 2018-687 of 3 August 2018, and by the Ordinance n°2018-1125 of 12 December 2018 (which will enter into force on 1 June 2019 at the latest).

Unlike the German legislator, who adopted a whole new statute to comply with the GDPR, the French chose to modify the "*Loi informatique et libertés*" (LIL) which was one of the first comprehensive data protection laws in Europe (dating back to 1978).

The processing of personal data for scientific and archiving purposes is regulated in articles 78 and 79 (according to the new numbering, which will enter into force on 1 June 2019). Article 78 provides that when data are processed for scientific purposes, certain rights of data subjects (access, rectification, restriction and the right to object) can be limited. The exact conditions in which such limitations are possible are to be specified by a Decree (*Décret en Conseil d'Etat*) which to the best of our knowledge has not yet been adopted.

Article 79 concerns purpose extension. It specifies that when data were collected for a different purpose and then re-used for research purposes (according to the purpose extension principle), the obligation to provide information to data subjects (art. 14 GDPR) does not apply.

Germany is probably the first country to have adopted a comprehensive national law to complete the General Data Protection Regulation. The new Bundesdatenschutzgesetz (BDSG) was adopted on June 30, 2017.

It shall be kept in mind that BDSG only applies to the processing of personal data by private entities and by public bodies of the German Federation (Art. 1 of the BDSG). Processing of personal data by public bodies of the Länder (such as universities) is governed by regional norms (Landesdatenschutzgesetze, LDSG). To the best of our knowledge, no LDSGs has yet been updated to conform to the GDPR. Therefore, for now, the situation regarding the processing of personal data for research purposes in German universities is not entirely clear.

As far as public bodies of the Federation (such as certain research institutes) and private entities are concerned, the processing of personal data for research purposes will be governed by Art. 89 of the GDPR, completed by section 27 of the new BDSG. The latter contains four paragraphs.

Firstly, section 27(1) of the new BDSG allows for processing of special categories of personal data for research purposes "if such processing is necessary for these purposes and the interests of the controller in processing substantially outweigh those of the data subject in not processing the data". The provision is based on Art. 9(2)(j) of the GDPR, which seems to leave the Member States the decision on whether to allow processing of special categories of data for research purposes based on the balance of interests. Interestingly, the new German law also contains a list of possible 'appropriate safeguards' for such processing[22]. The list is not meant to be exclusive, and other safeguards are also possible; moreover, it only expressly applies to the cases where special categories of data are processed. Moreover, as the GDPR does not expressly do it, section 27(3) of the new BDSG (still based on Art. 9(2)(j) of the GDPR) states that (according to the general principle of s. 89(1) of the GDPR) special categories of personal data processed for research purposes shall be pseudonymised, and then anonymised as soon as the purposes allow it.

[22] The safeguards "may include in particular the following: 1. technical organizational measures to ensure that processing complies with Regulation (EU) 2016/679; 2. measures to ensure that it is subsequently possible to verify and establish whether and by whom personal data were input, altered or removed; 3. measures to increase awareness of staff involved in processing operations; 4. designation of a data protection officer; 5. restrictions on access to personal data within the controller and by processors; 6. the pseudonymization of personal data; 7. the encryption of personal data; 8. measures to ensure the ability, confidentiality, integrity, availability and resilience of processing systems and services related to the processing of personal data, including the ability to rapidly restore availability and access in the event of a physical or technical incident; 9. a process for regularly testing, assessing and evaluating the effectiveness of technical and organizational measures for ensuring the security of the processing; 10. specific rules of procedure to ensure compliance with this Act and with Regulation (EU) 2016/679 in the event of transfer or processing for other purposes".

Secondly, section 27(2) provides for derogations from certain rights of data subjects, i.e., the right of access, rectification, restriction of processing and right to object. As suggested by Art. 89(2) of the GDPR the derogations apply when these rights are likely to render impossible or seriously impair the achievement of the research purposes and are necessary for their fulfilment. The German federal legislator has therefore taken full advantage of the leeway left by Art. 89(2) of the GDPR and legislated in favour of freedom of research.

Moreover, the legislator even went further than expressly allowed by this article and allowed for a derogation from the right of access when the provision of information listed in Art. 15(1) of the GDPR would involve a disproportionate effort. This derogation seems to be based on recital 62 of the GDPR.

Finally, section 27(4) of the new BDSG states that the controller may publish personal data (processed for research purposes) only if the data subject has provided consent or if doing so is indispensable for the presentation of research findings on contemporary events. This seems to serve as a limit to Art. 89 of the GDPR by stating that, in principle, special rules concerning research stop where publication of personal data starts.

In **Greece**, a Draft Bill for Personal Data (PDPA 2018c) implementing the GDPR after public consultation (which was completed on March 5, 2018), has been adopted and put into force as of 25 May 2018. The Bill contains an article dedicated to the processing of PD for "scientific or historical research or for statistical data". Processing of PD is allowed *if the subjects have given their consent for this or previous studies on the same data, if the data come from publicly accessible sources* or *if the processing can be proven to be required for the research*. For the processing of the special categories, the Bill is more restrictive; especially for research on genetic data prior consultation with the Data Protection Authority is mandatory. Medical data processing is allowed, provided the researchers involved are legally or professionally bound by confidentiality. Pseudonymisation or anonymisation are recommended but only when they do not hinder the purposes of the research. Overall, this draft Bill can be considered favourable towards research purposes.

The **Italian** Republic transposed the GDPR by legislative decree No 101/2018, which entered into force on 19 September 2018 (Italian law). According to that, personal data for scientific research can be processed without the consent of the data subject in the following cases: i) scientific research has been pursued according to the provision of law, provided that the data controller carries out an impact assessment and makes it publicly available, analysing the necessity and proportionality of the processing, the risks with respect to the rights and freedoms of data subjects, and safety measures to deal with these risks; ii) due to particular reasons, informing the data subject about the processing of personal data proves impossible or would involve a disproportionate effort, and it is likely to render impossible or seriously impair the achievement of the research objectives, provided that: a) the data controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, b) the research project has received favourable and motivated opinion from the Ethics Committee, c) the research project has been submitted to preventive consultation with the Italian Data Protection Authority (It. Garante per la protezione dei dati personali) and to an impact assessment, in accordance with Art. 35 and 36 of GDPR.

In accordance with Art. 110-bis of the Privacy code – as modified by legislative decree No 101/2018 – the reuse of data for research purposes is allowed when: i) it is carried out by third parties that mostly deal with research activities, ii) the information about the processing of personal data proves impossible or would involve a disproportionate effort, and it is likely to render impossible or seriously impair the achievement of the research objectives, iii) it is subject to prior authorization by the Italian Data Protection Authority, made dependent on the adoption of appropriate action in compliance with Art. 89 of GDPR. With specific reference to prior authorisation by the Italian Data Protection Authority, decisions on an application submitted in accordance with Art. 8 of legislative decree No 101/2018 shall be adopted and communicated to the applicant within 45 days after its receipt. The absence of delivery shall take the place of refusal. Also, Art. 8 of legislative decree No 101/2018 provides for the Italian Data Protection Authority to allow the reuse of data for research purposes also by means of general measures.

The Italian Data Protection Authority has been organising several information meetings with Italian universities and public research bodies to raise awareness among the different research communities

and university administrative staff on the changes introduced by the GDPR and their impact on research activities[23].

The next example is **Lithuania**. To duly comply with GDPR the new version of Lithuanian Law on Legal Protection of Personal Data (LLPPD 2018) was enacted and entered into force since 16 July 2018. The previous version of the law included a special exemption for scientific research in Art. 12, which contained quite detailed requirements for the procession of personal data without the data subject's consent. Among other things, the prior checking procedure by the State Data Protection Inspectorate was required. In contrast with the previous regulation and with Estonian and Finnish models as described above, the newly enacted LLPPD 2018 contains no special provisions dealing with the research exemption. The requirement of the prior checking procedure was abandoned as well. It means that Lithuania has not used the opportunities and flexibilities provided in Art. 89 of GDPR. It also means that after the implementation of GDPR, the persons using personal data for scientific research have to rely directly on and comply with the general provisions of GDPR, especially Art. 6, Art. 17.3 and Art. 89. Following the new regulation, Lithuanian universities and other research institutions have enacted their own internal rules, dealing, *inter alia*, with the research exception. For example, Vilnius University, which is the leading research and study institution in Lithuania, enacted the rules on the data protection, which prescribes, that university has a right to process personal data for scientific or historical research purposes. The same rules also state, that, in line with Art. 17.3 of GDPR, the right to be forgotten is not applicable when processing is necessary for, among others, scientific research purposes.

Since the legislative changes were enacted very recently, so far there are no reported cases of application or conflicts concerning the new regulation of research exception. Therefore, the real impact of GDPR on scientific research is yet to be seen.

5 Conclusion

The development and use of language resources often involve the processing of personal data. Several aspects of personal data may be confusing. For instance, it is arguable whether human voice as biometric data should be considered to belong to special categories of personal data (sensitive data). It should also be emphasised that publicly available data are protected by the GDPR as well.

The legal framework setting for requirements for processing personal data for research purposes is based on the GDPR and national laws of the EU Member States. This means that in addition to the GDPR, researchers that wish to develop and use LRs for language research must further follow national requirements.

References

[BDSG] Bundesdatenschutzgesetz. Available at https://www.gesetze-im-internet.de/bdsg_2018/index.html (5.9.2018)

[C-101/01] Case C-101/01. Criminal proceedings against Bodil Lindqvist (6 November 2003). Available at http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1521039149443&uri=CELEX:62001CJ0101 (3.4.2018)

[C-73/07] Case C-73/07. Tietosuojavaltuutettu vs. Satakunnan Markkinapörssi Oy and Satamedia Oy (16 December 2008). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62007CA0073&qid=1536154290371&from=EN (5.9.2018)

[Czech law] Zákon č. 101/2000 Sb. Zákon o ochraně osobních údajů a o změně některých zákonů. Available at https://zakonyprolidi.cz/cs/2000-101 (28.1.2019)

[Data Protection Directive] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L 281, 23/11/1995 p. 0031 – 0050. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046&qid=1522340616101&from=EN (29.3.2018)

[DPA 2018]. Data Protection Act (Finland). Entry into force 01.01.2019. Available in Swedish at: http://www.finlex.fi/sv/laki/alkup/2018/20181050 (20.1.2019)

[23] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/8318508 and https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/7977380 [accessed 22.03.2019]

[PDPA 2018a] Estonian Personal Data Protection Act (Isikuandmete kaitse seadus). Entry into force 15.01.2019. Available in Estonian at https://www.riigiteataja.ee/akt/104012019011 (21.1.2019)

[Draft PDPA 2018b] Finnish Draft Act on Personal Data Protection (Hallituksen esitys eduskunnalle EU:n yleistä tietosuoja-asetusta täydentäväksi lainsäädännöksi) (01.03.2018). Available at https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Sivut/HE_9+2018.aspx (4.4.2018)

[Draft PDPA 2018c] Greek Draft Bill on Personal Data Protection (Νόμος για την Προστασία Δεδομένων Προσωπικού Χαρακτήρα). Available at http://www.opengov.gr/ministryofjustice/wp-content/uploads/downloads/2018/02/sxedio_nomou_prostasia_pd.pdf (18.4.2018)

[French law] Loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles, modifying the French Data Protection Act (loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés)

[GDPR] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ L 119, 4.5.2016, p. 1-88. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1515793631105&uri=CELEX:32016R0679 (29.3.2018)

[González-Rodríguez et. al. 2008] Joaquín González-Rodríguez, Doroteo Torre Toledano, Javier Ortega-García (2008). Voice Biometrics. In Handbook of Biometrics edited by Anil K. Jain, Patrick Flynn, Arun A. Ross. Springer

[IPDPC] Italian Personal Data Protection Code. Legislative Decree 30.06.2003 No. 196. English version available at: http://194.242.234.211/documents/10160/2012405/Personal+Data+Protection+Code+-+Legislat.+Decree+no.196+of+30+June+2003.pdf (11.4.2018)

[Italian law] DECRETO LEGISLATIVO 10 agosto 2018, n. 101 Disposizioni per l'adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonche' alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati).(18G00129) (GU n.205 del 4-9-2018). The Italian version of the law available at http://www.gazzettaufficiale.it/eli/id/2018/09/04/18G00129/sg (27.1.2019).

[Jain et. al. 2004] Anil K. Jain, Arun Ross, Salil Prabhakar (2004). An Introduction to Biometric Recognition. - IEEE Transactions on Circuits and Systems for Video Technology 14(1). Available at https://www.cse.msu.edu/~rossarun/BiometricsTextBook/Papers/Introduction/JainRossPrabhakar_BiometricIntro_CSVT04.pdf (31.3.2018)

[Kelli et al. 2015] Aleksei Kelli, Kadri Vider, Krister Lindén (2015). The Regulatory and Contractual Framework as an Integral Part of the CLARIN Infrastructure. 123: Selected Papers from the CLARIN Annual Conference 2015, October 14–16, 2015, Wroclaw, Poland. Ed. Koenraad De Smedt. Linköping University Electronic Press, Linköpings universitet, 13−24. Available at http://www.ep.liu.se/ecp/article.asp?issue=123&article=002 (28.3.2018)

[LED] Law of European Delegation. Law No. 25.10.2017 No 163. Available at http://www.gazzettaufficiale.it/eli/id/2017/11/6/17G00177/sg (11.4.2018)

[LLPPD 2018] Lithuanian Law Amending the Law on Legal Protection of Personal Data (Lietuvos Respublikos asmens duomenų teisinės apsaugos įstatymo pakeitimo įstatymas). Available at https://www.e-tar.lt/portal/legalAct.html?documentId=43cddd8084cc11e8ae2bfd1913d66d57 (30.8.2018)

[Oostveen 2016] Manon Oostveen (2016). Identifiability and the applicability of data protection to big data. International Data Privacy Law 6 (4), 299-309

[ORDA] Organisation of Research and Development Act. Entry into force 2.05.1997. English translation available at https://www.riigiteataja.ee/en/eli/513042015012/consolide (21.1.2019)

[Privacy Code] Code of conduct and professional practice Regarding the processing of personal data for historical purposes. English version available at http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/export/1565819 (11.4.2018)

[VLO] CLARIN Virtual Language Observatory. Available at https://vlo.clarin.eu/ (18.4.2018)

[WP29 2017] WP29. Guidelines on Consent under Regulation 2016/679. Adopted on 28 November 2017 [adopted, but still to be finalized]. Available at http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=615239 (2.4.2018)

[WP29 2014a] WP29. Opinion 05/2014 on Anonymisation Techniques Adopted on 10 April 2014. Available at http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf (3.4.2018)

[WP29 2014b] WP29. Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC. Available at http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf (3.4.2018)

[WP29 2007] WP29. Opinion 4/2007 on the concept of personal data. Adopted on 20th June. Available at http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf (29.3.2018)